

# A STUDY OF LONG-TERM STABILITY OF ATOMIC CLOCKS

David W. Allan  
National Bureau of Standards  
Time and Frequency Division  
325 Broadway  
Boulder, Colorado 80303

## ABSTRACT

The importance of long-term frequency stability has increased significantly in recent years because of the discovery of the millisecond pulsar, PSR 1937+21. In addition, long-term stability is extremely useful not only in evaluating primary frequency standards and the performance of national timing centers, but also in addressing questions regarding autonomy for the Global Positioning System (GPS) and syntonization for Jet Propulsion Laboratory's (JPL) Deep Space Tracking Network.

Over the last year, NBS has carried out several studies addressing questions regarding the long-term frequency stability of different kinds of atomic clocks as well as of principal timing centers. These analyses cover commercial and primary cesium-beam frequency standards; active hydrogen maser frequency standards and the new commercial mercury-ion frequency standards at United States Naval Observatory (USNO).

The support and cooperation of the timing centers involved have, along with the GPS common-view measurement technique, made these analyses possible. Fortunately, the measurement noise was significantly less than the clock noise for the long-term stability regions studied. Fractional frequency stabilities,  $\sigma_y(\tau)$ , of parts in  $10^{14}$  down to parts in  $10^{15}$  were observed for various of these standards. It is believed that frequency stabilities on the order of a part in  $10^{15}$  will be necessary to measure the effects of gravity waves on frequency stability between millisecond pulsars and atomic clocks on or near the earth.

## INTRODUCTION

The primary reasons for improved long-term stability in atomic clocks have been the study of millisecond pulsars<sup>[1,2]</sup>, autonomy questions for GPS, maintaining syntonization for the JPL Deep Space Network and long term stability performance of new clocks at the primary timing centers throughout the world. As needs have arisen for improved stability in atomic time scales, several laboratories have responded with very strong efforts in producing new standards<sup>[3-8]</sup>. Some analysis techniques to ascertain current performance of the best operating clocks in the world have also been developed. Fortunately, the GPS common-view time-difference and frequency-difference measurement technique has allowed us to analyze long-term clock time deviations at state-of-the-art levels.

One analysis tool used was a statistical cross-correlation technique in the time domain in order to ascertain correlations at different integration times<sup>[1]</sup>. This tool was developed to study long term correlations in the presence of nonstationary processes in the clock deviations. The next tool was to employ the NBS algorithm for generating the time scale NBS(AT1) from the ensemble of clocks at NBS<sup>[9]</sup>. NBS(AT1) plus an equation generates UTC(NBS); the latter being the official time from the National Bureau of Standards. This same algorithm was used on the international set of clocks available through GPS common-view measurements. Each laboratory timing center's output was treated as an

independent clock and then combined with all the others in an optimum minimum squared error sense according to the NBS algorithm's design. The hope is to generate a "world's best" clock, which in turn could be used to analyze each of the constituent members and then better understand the long term stability performance of each. The next tool used was an  $N$ -cornered hat statistical approach under the assumption of independence of the member clocks<sup>[10]</sup>. In this way we can sort out the performance of each member of a set of  $N$  clocks in terms of its statistical characteristics. Last, a second-difference drift estimator was used to study long term frequency drift correlations among some of the clocks involved<sup>[10,11]</sup>.

Figure 1 shows a plot of the fractional frequency stability between the millisecond pulsar (PSR 1937+21) and UTC(NBS), as well as the stability estimate of NBS(AT1) versus the "world's best" clock estimate. Figure 1 shows that with more integration time the millisecond pulsar might show better performance than the best clocks in the world. As indicated in the plot the possibility of measuring gravity waves at a few parts in  $10^{15}$  at integration times of the order of a year is by itself a significant motivator.

#### GENERATION OF THE "WORLD'S BEST" CLOCK AND COMPARISON OF THE PERFORMANCE OF THE INTERNATIONAL TIMING CENTERS

The advantage of post-analysis is that we can more easily discern outliers and abnormal performance in the member clocks. Following this post analysis and removal of outliers and abnormal behavior, we can then employ optimal statistical processing techniques. These processing techniques can approach the ideal once the nonstochastic behavior has been filtered off. In contrast, the time from any timing center, including the generation of TAI has to be calculated or generated in near real time. In a real-time operation, subtle frequency steps and annual variations cannot readily be dealt with because of their low frequency character. Often they are best observed in post-analysis.

A period of time commensurate with the installation of the new measurement system for the millisecond pulsar (PSR1937+21) at Arecibo, Puerto Rico was picked. The time was a 1110 day period from MJD 45769 to MJD 46879. The data not available on the NBS computer file were accrued from the Princeton University computer file or were extracted from the BIH annual report. The UTC scales were used from the various timing centers as they were more readily available. Since some of the UTC scales are steered, we tried to remove the steering corrections so that the contributing scales would be independent. The post-analysis was performed removing known frequency steps in the time scales and outliers that would contaminate the stochastic analysis.

An  $N$ -cornered hat analysis between the times generated by the various timing centers was first performed in order to get a first order estimate of the 10 day stability and the longer term stability of each of the outputs from the timing centers. Preliminary parameters were then chosen for the optimization routine for the NBS algorithm to include each timing center's output as a clock member. One of the features of the NBS algorithm is that it is an adaptive filter so that as a clock improves or degrades with time the filtered estimate will adapt to the value commensurate with the clock's performance. The time constant of this filter for this experiment was set at three months. Once the parameters have been set and the clocks processed through the NBS algorithm, the "world's best" estimate could then be used to better ascertain the performance of each member clock, and the parameters could be appropriately fine tuned to reduce the minimum squared error of the estimate of the ensemble as well as of each individual member. Figures 2a and 2b through 14a and 14b list a modified  $\sigma_y(\tau)$  fractional frequency stability plot and a time residual plot with a mean frequency and a mean time removed for each of the contributing timing centers. The abscissas and ordinates have been kept the same in the figures for convenience of comparison.

The contributing timing centers were: Applied Physics Laboratory at Johns Hopkins University, Laurel, MD; Commonwealth Scientific and Industrial Research Organization, Australia; Istituto Elettrotecnico Nazionale, Torino, Italy; National Bureau of Standards, Boulder, CO; National Physical Laboratory, Teddington, England; National Research Council in Ottawa, Canada; Paris Observatory, Paris, France; Physikalisch-Technische Bundesanstalt, Braunschweig, Germany; Radio Research Laboratory Tokyo, Japan; and Tokyo Astronomical Observatory, Tokyo, Japan; Technische Universität Graz, Graz, Austria; US Naval Observatory, Washington, DC; and Van Swinden Laboratory, Delft, Netherlands.

Following are some highlights to note from the several figures. Though there was a limited set of data available on the hydrogen maser at APL, the stability at integration times of a few months is very impressive. The NBS algorithm generates from its own ensemble a scale called AT1, which is neither synchronized nor syntonized with any other scale. NBS(AT1) is optimized for frequency stability. Because NBS(AT1) had the largest weight among the set of clocks used in this "world's best" estimate, we removed it from the ensemble and looked at it independently against all the other optimally weighted clocks in order to see whether a similar stability performance was still obtained. This independent analysis confirmed the stability plots shown in Figures 5a and 5b.

Most of the timing centers have clock output stabilities at levels of a few parts in  $10^{14}$  for integration times from 10 days out to the order of a year. Figures 9a and 9b for PTB show exceedingly good stability from year to year, though there may be inadequate confidence of the estimate to confirm this. This stability may be because UTC(PTB) is steered with their primary cesium clock. Also annual variations appear in the time deviations of the PTB output. The excellent performance of the TUG clock may be because this single commercial clock is temperature controlled, humidity controlled and voltage controlled. The apparent quadratic behavior observed in the residuals of the USNO clock is of concern because most of these clocks are high performance cesium beam clocks. We are led to ask whether there is a systematic frequency drift in a particular production line series. On the other hand, the month-to-month stability is very good.

#### A STUDY OF THE ANNUAL VARIATIONS

For the most stringent of timing users, B. Guinot at the International Bureau of Weights and Measures (BIPM) has generated a scale called Terrestrial Time BIPM TT(BIPM). Figure 15 shows the fractional frequency differences between TT(BIPM) and International Atomic Time (TAI) over the last several years. The noticeable frequency step is associated with the 1972 adjustment in the SI second used for TAI to conform with the NBS, NRC and PTB primary frequency standards. Annual variations are apparent between these two scales. Since PTB Cs 1 is believed to be one of the best primary clocks in the world, its weight in generating TT(BIPM) is quite high, and the annual term shown earlier in the PTB plot in Figures 9a and 9b corresponds with the annual variations shown in Figure 15. We are led to the questions: are these annual variations in all of the clocks in the world ensemble, moving up and down together on an annual basis, or are they in PTB Cs 1, or a combination of both? At the few parts in  $10^{14}$  level it is very difficult to ascertain which conclusion is the most nearly correct.

In an effort to obtain an independent estimate regarding this problem, the data from the millisecond pulsar were analyzed. The pulsar data have much larger variations from week to week and month to month — primarily due to the 4 k parsec (13,000 light-years) path of the signal. The variations can be fairly well characterized as a white noise phase modulation (PM) process at about the 300 ns level. Because such noise integrates as  $\tau^{-3/2}$  on a mod  $\sigma_y(\tau)$  plot, the long term variations can be meaningful. PSR1937+21 is plotted against PTB Cs 1 in Figure 16. Also plotted in Figure 16 are the time difference residuals between PTB Cs 1 and NBS(AT1). Visually, correlations are apparent even though masked

somewhat by the pulsar data noise.

To further study the significance of these annual variations we incorporated a cross-correlation analysis tool<sup>[1]</sup>. The correlations seemed statistically significant for integration times longer than several months. From this study, the annual variations seem to be primarily in PTB Cs 1; however, this is inconclusive because in the generation of the pulsar data, an annual term had to be removed in conjunction with the determination of the ephemeris for the earth as it orbits the sun. Further study is needed before a definite conclusion can be obtained.

## STUDY OF THE COMMERCIAL MERCURY ION STANDARD'S FREQUENCY STABILITIES

Three independent clocks were chosen to study the stability of the two commercial mercury-ion standards located at USNO <sup>[6]</sup>. Data on the unit called No. 2 covered about a 1 year period and the data on No. 3 covered the last one-half year period. As can be seen in the fractional frequency stability plots shown in Figures 17a, 18a and 19a, there are missing data and some abnormal behavior during the first several weeks of the data. This first part was segmented off for the statistical analysis period, which was from 19 December 1986 to 14 November 1987. The stabilities plotted in Figures 17b, 18b and 19b are for the segmented period.

Assuming independence, the four clocks (unsteered UTC(USNO), NBS(AT1), the APL hydrogen maser and the mercury ion No. 2) were compared in an *N*-cornered hat analysis. That data are plotted in Figure 20.

For integration times of one, two and four days the GPS common-view noise contaminated the data for APL and NBS. It appears that the NBS and USNO outputs have a flicker frequency modulation (FM) like performance at about a part in  $10^{14}$  over this 11 month data set. The long term frequency drift in the APL hydrogen maser is apparent. The excellent stability performance of the mercury ion standard — as the winner in this 4-cornered hat analysis for integration times of a week to a month — is also apparent. The one, two and four day stability values plotted may be pessimistic, as they are contaminated with measurement noise. However, there appears to be some stability degeneration for integration times of the order of a few months and longer. For completeness the other pair-wise members of the 4-cornered hat are shown in Figures 21a and b through 23a and b.

Over the last one-half year we studied the stability difference between the two commercial mercury-ion standards No. 2 and No. 3. Figures 24a and b show a most outstanding performance of these two standards. Since the stability plotted in Figure 24 is inconsistent with that plotted in Figure 20, we asked whether No. 1 and No. 2 were correlated. To study whether these two clocks have long term correlations and to be able to define a cross-correlation coefficient as a function of  $\tau$  called  $\rho(\tau)$  we proceeded as follows<sup>[10]</sup>. Given fractional frequency differences:

$$y_1 = \text{clock i} - \text{clock j and}$$

$$y_2 = \text{clock k} - \text{clock l,}$$

we can compute a variance of the difference

$$\sigma_{y_1 - y_2}^2(\tau) = \sigma_{y_1}^2(\tau) - 2\sigma_{y_1 y_2}(\tau) + \sigma_{y_2}^2(\tau)$$

We construct:

$$\rho(\tau) = \frac{\sigma_{y_1 y_2}(\tau)}{\sigma_{y_1}(\tau) \cdot \sigma_{y_2}(\tau)}.$$

At a given  $\tau$ , if  $\rho(\tau)$  approached +1 then we could conclude that there are correlations between  $i$  and  $k$ , or  $j$  and  $l$ , or  $-i$  and  $l$ , or  $-k$  and  $j$ . If  $\rho(\tau)$  approached -1, then we could conclude that there are correlations between  $-i$  and  $k$ , or  $-j$  and  $l$ , or  $i$  and  $l$ , or  $k$  and  $j$ . This cross-correlation analysis was tested with simulated data and gave affirmative results to the above conclusions and to the ability to test in the presence of low frequency divergence processes such as flicker noise FM and random walk FM. Figure 25 would indicate that some of the long-term instabilities are in fact due to mercury-ion No. 2. Figure 26a and b would indicate that No. 2 and No. 3 are positively correlated for integration times of a few months. Using different references, Figures 27a and b lead us to the same conclusion as similarly do Figures 28a and b. All three independent clock pairs lead to the same conclusion that mercury-ion standards No. 2 and No. 3 appear to be long term correlated for integration times of a few months. Figure 29a would indicate no statistical correlation between the four clocks involved. On the other hand Figure 29b shows a definite positive cross-correlation; this is believed to be due to the fact that over the half year analyzed mercury-ion standard No. 2 and the APL maser had positive frequency drifts.

The following independent test of the hypotheses that the low frequency variations could be modelled by long-term correlations between the two mercury ion standards and frequency drifts between the APL maser and the mercury ion standards, was conducted. Frequency drifts were measured using a second difference estimator over the last half year with respect to NBS(AT1) UTC(USNO-unsteered) and the UTC(PTB) time scales. UTC(PTB) has as its reference the PTB primary Cs 1 standard. We observe a very high correlation in the frequency drift numbers measured with respect to the different laboratories (Figure 30). Within this small sample of three clocks, the drifts seem to be statistically significant — indicating, again, that in long term the two mercury ion standards appear to be correlated, and that the frequency drifts of the APL hydrogen maser and the ion standards are in the same direction.

The correlation mechanism is not yet understood and is being studied. Environmental perturbations due to temperature, temperature gradients and pressure are possible candidates. The very outstanding performance of these mercury ion standards gives great hope that if this problem can be identified and fixed, these standards will have a major contribution in the long-term stability analysis of the best frequency standards known.

## CONCLUSIONS

The long term frequency stability of time scales has improved dramatically over the last few years. This study has revealed that additional long term frequency stability improvements may be possible if the cause of annual variations in clocks can be identified and fixed. Also, if frequency drift problems can be solved, and if cross-correlation problems through the environment can be identified and fixed, the next few years should show significant improvements in the performance of hydrogen masers and of mercury-ion frequency standards.

It is also apparent from this study that algorithms for combining clock readings can significantly impact the performance of a computed time scale. The NBS algorithm implemented seemed to have merit even where there were significant differences in statistical characteristics of the member clocks contributing to the ensemble.

## ACKNOWLEDGEMENTS

The author is greatly indebted to a large number of people. The cooperation from the International Timing Centers has been outstanding. Specifically, the USNO staff has been most cooperative in providing their data. The author is greatly indebted to Trudi Pepler